Computational Statistics and Data Analysis 131 (2019) 176–193

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Quantifying the risk of heat waves using extreme value theory and spatio-temporal functional data

Joshua French [a], Piotr Kokoszka [b,*], Stilian Stoev [c], Lauren Hall [a]

[a] University of Colorado, United States
[b] Colorado State University, United States
[c] University of Michigan, United States

ARTICLE INFO

Article history:
Received 6 March 2018
Received in revised form 10 July 2018
Accepted 11 July 2018
Available online 20 July 2018

Keywords:
Extreme events
Functional data
Elsevier
Heat waves
Spatio-temporal analysis

ABSTRACT

Heat waves and other extreme weather events have attracted a great deal of attention due to their socioeconomic impacts and relation to climate change. A heat wave is defined through a general loss function that captures its amplitude, temporal persistence, and spatial extent. The proposed statistical framework is at the nexus of extreme value theory (EVT) and functional data analysis (FDA) and enables computation of probabilities of yet unobserved rare events that are not seen in historical records. Data from the North American Regional Climate Change Assessment Program, which has produced computer model predictions of current and future temperatures across much of North America, are used. The approach allows for the computation of probabilities for heat waves of any pre-specified temporal duration, spatial extent, and overall magnitude. It can be applied to the computation of probabilities of other extreme weather events, including cold spells and droughts.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

The purpose of the research reported in this paper is to propose a flexible and readily applicable framework that allows us to compute probabilities of waves of extreme weather. While we focus on heat waves, our approach can be applied to the computation of probabilities of other extreme weather events, including cold spells and droughts. Our work is to a large extent motivated by data from the North American Regional Climate Change Assessment Program (NARCCAP; Mearns et al., 2007, updated 2014), which has produced computer model predictions of current and future temperatures across much of North America. Our approach thus contains two types of prediction: (1) those implied by the NARCCAP computer model, and (2) those obtained by the application of extreme value theory spatio-temporal functional data. It allows us to compute probabilities of heat waves of any prespecified temporal duration, spatial extent and intensity.

There is no universally agreed definition of a heat wave. Heat wave definitions depend on the objective of a study; relevant research is reviewed in the Supplemental Material. However, all definitions combine temporal duration and some threshold levels. If a study extends beyond a specific location or a relatively small region, the spatial aspect of the data is also incorporated. For smaller regions, definitions based on fixed thresholds or historical quantiles are useful, especially from the public health point of view. However, such definitions may not be relevant to environmental studies, as plant and animal species adapted to specific regions respond to temperature changes far above (or below) what is typical for the environment they inhabit rather than to those crossing fixed thresholds. In this respect, measures based on locally-defined percentiles are

* Corresponding author.
E-mail address: Piotr.Kokoszka@colostate.edu (P. Kokoszka).

https://doi.org/10.1016/j.csda.2018.07.004
0167-9473/© 2018 Elsevier B.V. All rights reserved.

more appropriate. However, a limitation of such an approach is that upper percentiles reflect historical records, so it is not clear how to use them to compute probabilities of heat waves whose amplitude, duration, and spatial extent are not seen in historical records. Since we aim at developing a tool set for computing probabilities of possibly yet unobserved rare events, we propose a different definition. It quantifies departures from values typical for a specific region over a specific period of time, with the magnitude of the departure being potentially extremely high. As with the other studies, the definition we propose is motivated by the objective we want to achieve and the data we use. No claim of its universal superiority is made.

As noted above, our objective is to compute the probability of an extreme heat wave of a prespecified duration, amplitude, and spatial extent. Many previous studies (see the Supplemental Material) use specific probabilistic models, typically involving some hierarchy with point process and continuous distributions components. Since we focus on extreme events, we do not need to specify a detailed probabilistic mechanism that heat waves occurrences should follow. We instead use their limit extremal behavior, which is justified by the heat wave definition we propose. This relatively simple approach is suitable for our purpose. As always, some precision can be lost due to the application of limit results, but biases resulting from a misspecified probabilistic model can be reduced. Additional advantages of the proposed methodology are that it is fast, simple, and can produce easily interpretable and practically useful results. It is hoped that it will complement useful research that has recently been done.

The paper is organized as follows. We introduce the NARCCAP Data in Section 2. Section 3 focuses on the definition of a heat wave and shows how it allows us to identify regions at risk of extreme heat waves. We derive a general loss function that can be used to quantify heat waves, along with methodology based on extreme value theory that allows us to compute the probability of a heat wave. The NARCCAP data are analyzed in Section 4. Section 5 presents the results of several simulation studies designed to assess the performance of our approach. The contribution of the paper is summarized in Section 6.

2. NARCCAP data

The overarching goal of the NARCCAP is to explore the uncertainty associated with using atmosphere–ocean general circulation models (GCMs) to drive regional climate models (RCMs) that downscale climate to a finer resolution. We utilize data from the second phase of the program, in which four GCMs provide boundary conditions for six RCMs during a future period (2041–2070). See Mearns et al. (2009) for more details about the experimental design and climate conditions.

We specifically examine output from the Canadian Regional Climate Model (CRCM; Caya and Laprise, 1999) combined with the Community Climate System Model (CCSM) global climate model (Collins et al., 2006). The variable of interest is maximum daily surface air temperature. The temperature values are available at 16,100 sites on a 140×115 grid covering much of North America, the grid area is shown in Fig. 1, in Section 3. The model output includes daily values over nearly 30 years during the period 2041–2070. Because of incomplete records in the final year, only the first 29 years of data were utilized in analysis.

3. Heat waves: definition and risk quantification

This section develops the statistical framework we use in Section 4 to analyze the NARCCAP data. The framework is general and can be applied to other similar data sets and to different extreme weather events characterized by amplitude (intensity), temporal duration, and spatial extent. We discuss these three characteristics in Section 3.1 in the context of a heat wave. In Section 3.2, we quantify them by means of suitable loss functionals. Section 3.3 develops an approach that allows us to compute probabilities of extreme heat waves. It is based on the application of EVT to functionals computed from annual temperatures curves.

3.1. Components of a heat wave

This section motivates the definitions we introduce in Section 3.2. For illustration, we use the NARCCAP data described in Section 2.

A heat wave definition should be able to capture features related to, "... producing regional impacts on environment and society, including health, infrastructure, and economy" (Gershunov et al., 2009). There is a broad agreement that a heat wave's overall size or *magnitude* can be captured through the following three components:

- *amplitude (or intensity)*: the size of the exceedance of the temperature over some reference threshold.
- *duration*: the amount of time over which the heat wave persists.
- *spatial extent*: the geographic area covered by the heat wave.

The threshold used for quantifying the amplitude can be *absolute* or *relative*. An example of an absolute threshold is 105 °F (40.5 °C). Such a threshold is motivated by public health considerations, but it is not suitable to study the environmental impact of heat waves in various parts of the Earth. For example, it is not relevant to the evaluation of the probability of heat waves in Antarctica, where the highest temperature ever recorded is only slightly above 60 °F (15.6 °C) (Burt, 2015). Since our objective is to compute probabilities of heat waves over a large region (North America) with climates ranging from subtropical to arctic, an absolute threshold cannot be used. One might instead use a fixed relative threshold, such as 5 °C above the historical mean. However, the variability of temperature can vary greatly across a large spatial domain.

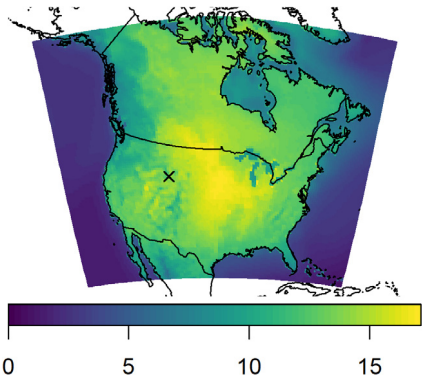

Fig. 1. Heat map of the standard deviation of daily temperature measurements (C) calculated over time at each point across the NARCCAP domain. The × marks the site in Wyoming where temperatures (and related functions) are shown over time in many subsequent figures.

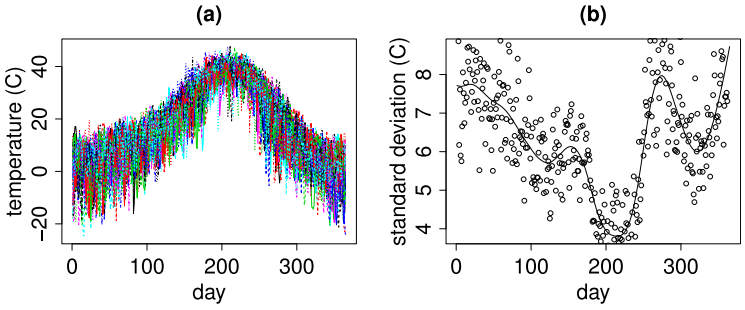

Fig. 2. Time series plots for a location in Wyoming. Panel (a) shows temperature (C) time series for each year between 2041 and 2069. Panel (b) shows the time series of standard deviation of daily temperature (C) across year with a smoothing spline overlaid.

Fig. 1 displays the standard deviation of daily temperatures for the NARCCAP data as a function of location. These standard deviations, which roughly reflect the difference between the minimum nighttime and maximum daytime temperatures, vary from about 7 °C in coastal areas to about 17 °C over the Great Plains. Additionally, the variability of the temperature at a fixed location generally depends on the time of the year. Fig. 2(a) overlays numerous annual daily temperature time series for a location in Wyoming (see the "×" in Fig. 1), while (b) depicts the standard deviation of daily temperature (across years) as a function of day. Generally, in North America, the variability of temperature is higher during winter months than summer months. So in addition to accounting for the heteroscedasticity of variances across the spatial domain, one should also account for the seasonal heteroscedasticity. Thus, when quantifying amplitude, it makes sense to either standardize the temperatures (so that thresholds are related to standard deviations above the mean) or use quantiles as thresholds that vary over time and space. Since empirical quantiles do not allow us to go beyond the range of the observed data, we adopt an approach based on standardized temperatures in most of what follows.

The duration of a heat wave can be quantified in various ways, but two are commonly used. The first is by taking an ℓ-day moving average of temperatures or their exceedances over a threshold; the values for the current and $\ell - 1$ previous days are averaged in some way, e.g. Hajat et al. (2002). One could instead compute an ℓ-day running median. The second approach for capturing the duration of a heat wave is to compute the ℓ-day running minimum of the exceedance amplitude or temperatures, Beniston (2004). Each of these methods describes the persistence of recent temperature values over time and thus captures the *duration* component of a heat wave. In most public health applications the duration ℓ is set to be fairly small (2 or 3 days). However, the impacts of heat waves on agriculture and economy are better quantified using larger values of ℓ (e.g., 7 or 14 days) combined perhaps with lower exceedance thresholds. For example, if the temperature exceeds 5 °F (2.8 °C) over the mean seasonal value in the summer for a period of 2 weeks, this can result in severe stress on the irrigation infrastructure, depletion of resources, and crop-loss.

Lastly, we discuss the spatial extent of a heat wave. Gershunov et al. (2009) define spatial extent as the percentage of representative stations where local thresholds are exceeded. Such an approach is suitable to study the extent of a specific historical heat wave. Our objective is to compute the probabilities of possible future heat waves and identify regions at risk. We must therefore first specify the spatial size of heat waves of interest; spatially large heat waves will appear with smaller probabilities than small localized heat waves. We propose two measures for quantifying the spatial extent of a heat wave. The first is to consider data within a distance d of a reference location. The second, to consider data for the k nearest neighbors.

For gridded data, like the NARCCAP data, these approaches are nearly equivalent. In principle, the method could be applied to neighborhoods consisting of a single point, for example if a specific city is of interest due to public health concerns.

In the next section, we formally describe the methods for quantifying the three components of a heat wave, and then combine them into a unified loss function quantifying the magnitude of a heat wave.

3.2. Quantifying the magnitude of a heat wave

Data representation

The raw data we consider are spatially-indexed time series of daily temperature measurements:

$$X(\mathbf{s}_i, j),\ j = 1, \ldots, T, \tag{1}$$

denoting the temperature at a spatial location $\mathbf{s}_i \in \mathbb{R}^2$, $i = 1, \ldots, m$, on day $j = 1, \ldots, T$. Due to the natural annual climate cycle, for each site, we partition $\{X(\mathbf{s}_i, j)\}$ into years and view the resulting 365-dimensional vectors as samples from a *functional time series*:

$$X_n(\mathbf{s}_i, \cdot) = (X_n(\mathbf{s}_i, t),\ t \in [0, 1]). \tag{2}$$

Here, $t \mapsto X_n(\mathbf{s}_i, t)$ is the temperature curve at site $\mathbf{s}_i$ for year n, viewed as a function of time t (measured as a fraction of the year). Since we consider daily temperatures (either daily average, minima, or maxima), the functional time series is *sampled* at times $t_k = k/365$, with $k = 1, 2, \ldots, 365$. The raw daily time series (1) and the functional time series (2) are related via the formula

$$j = 365(n_j - 1) + k_j,\ 1 \le k_j \le 365,\ n_j = 1, \ldots, N, \tag{3}$$

where n_j is the year of daily measurement j, k_j is the day within year n_j corresponding to the jth daily measurement, and $[\cdot]$ denotes the integer part of a real number. Therefore,

$$X_{n_j}(\mathbf{s}_i, t_{k_j}) \equiv X(\mathbf{s}_i, j), \quad \text{where } t_{k_j} := k_j/365.$$

For example, $X_3(\mathbf{s}_4, 33/365) = X(\mathbf{s}_4, 1128)$ is the temperature at location 4 on February 2 of the third year in the sample. Leap years are ignored in the simulated NARCCAP data, and all years have 365 days. For further reference, Fig. 3(a) shows the raw daily temperature (C) time series at a site in Wyoming for a period of four years, while Fig. 2(a) shows the corresponding realizations of the annual (sampled) functional time series curves, $X_n(\mathbf{s}_i, \cdot)$.

In the following subsections, we introduce loss functionals which allows us to quantify the *amplitude (intensity)*, *duration*, and *spatial extent* of heat waves. Our approach to the quantification of the amplitude is based on standardization.

Data standardization

One way to incorporate spatial as well as seasonal variability is to use standardized temperature values:

$$Z_n(\mathbf{s}_i, t_k) = \frac{X_n(\mathbf{s}_i, t_k) - \overline{X}(\mathbf{s}_i, t_k)}{\mathrm{SD}(\mathbf{s}_i, t_k)}, \tag{4}$$

where

$$\overline{X}(\mathbf{s}_i, t_k) = \frac{1}{N}\sum_{n=1}^{N} X_n(\mathbf{s}_i, t_k) \ \text{ and } \ \mathrm{SD}^2(\mathbf{s}_i, t_j) = \frac{1}{N-1}\sum_{n=1}^{N}(X_n(\mathbf{s}_i, t_k) - \overline{X}(\mathbf{s}_i, t_k))^2 \tag{5}$$

define the sample mean and sample variance, respectively, of the temperatures at site $\mathbf{s}_i$ for day t_k across the sample of N years. The $Z_n(\mathbf{s}_i, t_k)$ are unitless and can be interpreted as standard deviations of the temperature from the mean at location $\mathbf{s}_i$ and time $t_k \in [0, 1]$ of year n. Alternative measures of center and spread such as medians and inter-quartile ranges could be used in place of the sample mean and standard deviation, respectively. The ultimate goal of standardization is to make the methodology adaptive to the local level of temperatures over the seasons and across space. The time series

$$Z(\mathbf{s}_i, j) \equiv Z_{n_j}(\mathbf{s}_i, t_j),\ \ j = 1, 2, \ldots, T = 365 \times N, \tag{6}$$

is obtained by concatenating the yearly temperature curves of the standardized functional time series data. Fig. 3(a) shows the raw temperature time series for the first four years of NARCCAP data at the site in Wyoming depicted in Fig. 1, while (b) shows the standardized temperatures. Observe that the non-stationarity in the mean and variance of the process over time has been attenuated through standardization. Fig. 4 shows autocorrelation function (ACF) plots for the full $N = 29$ years of raw and standardized daily temperature data for the site in Wyoming. The exponential decal of the ACF plot in panel (b) suggests that the series is at least weakly stationary. Standardization thus has two practical benefits: 1) we can work with an absolute threshold, 2) the time series $\{Z(\mathbf{s}_i, j),\ j = 1, 2, \ldots, T\}$ in (6) can be assumed to be stationary. These two properties will allow us to apply EVT methods, which require a fixed threshold and stationary observations.

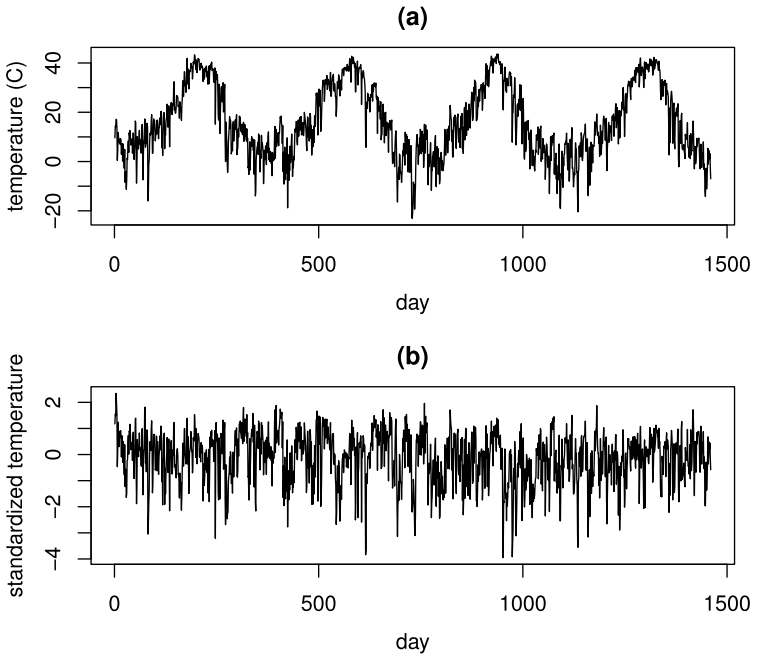

Fig. 3. Time series plots for a site in Wyoming over 4 years. (a) depicts raw daily temperature (C) and (b) depicts standardized temperature.

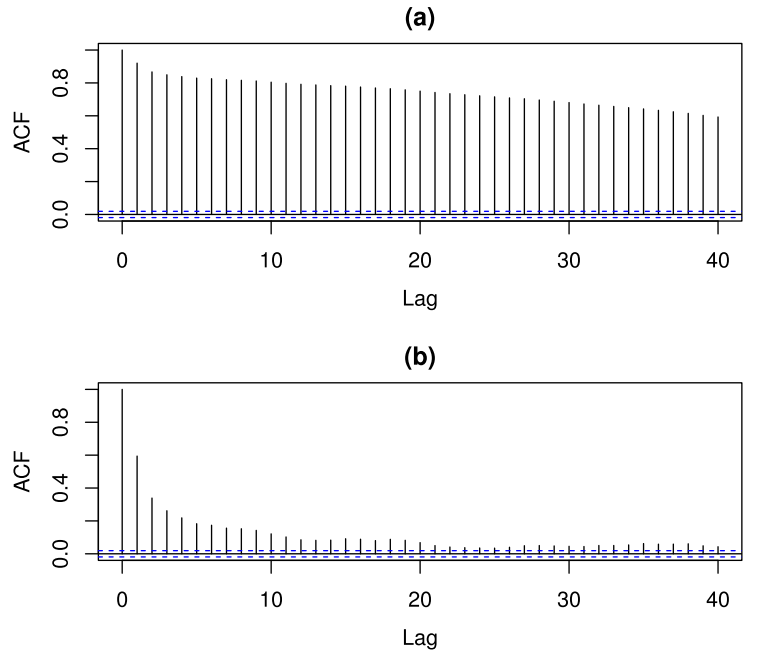

Fig. 4. ACF plots for a site in Wyoming over 29 years. (a) ACF for raw daily temperature and (b) the ACF for the standardized daily temperature.

Amplitude threshold

Let $Y(\mathbf{s}_i, t),\ t \in [0, 1]$ be a collection of spatially-indexed functional time series curves sampled at $t_k = k/365,\ k = 1, 2, \ldots, 365$. For example, the $Y(\mathbf{s}_i, \cdot)$ can be raw annual curves $X_n(\mathbf{s}_i, \cdot)$ or the standardized curves $Z_n(\mathbf{s}_i, \cdot)$. The amplitude of a heat wave is characterized by the relationship between Y and some temperature threshold u. The threshold u is chosen to refer to some amplitude of interest. If we were interested in quantifying heat waves hotter than 30 °C for the raw temperature

data, then $u = 30\,^{\circ}$C. Alternatively, if Y were actually the standardized temperature process Z, then one might be interested in assessing heat waves more than 2 standard deviations above the mean, in which case $u = 2$.

Duration functionals

To capture the persistence of a heat wave, we introduce a functional $\mathcal{D}_\ell$ that involves a duration parameter ℓ. One possible such functional is the temporal moving average of Y, that is,

$$\mathcal{D}_\ell(Y)(\mathbf{s}_i, t_k) = \mathcal{D}_\ell^{\text{avg}}(Y)(\mathbf{s}_i, t_k) := \frac{1}{\ell} \sum_{l=k-\ell+1}^{k} Y(\mathbf{s}_i, t_l). \tag{7}$$

An alternative function involves the minimum of Y over the time interval of interest, defined as follows

$$\mathcal{D}_\ell(Y)(\mathbf{s}_i, t_k) = \mathcal{D}_\ell^{\min}(Y)(\mathbf{s}_i, t_k) := \min_{l=k-\ell+1,\ldots,k} Y(\mathbf{s}_i, t_l). \tag{8}$$

When applied to the raw temperature data, $\mathcal{D}_\ell^{\min}(X_n)(\mathbf{s}_i, t_k) > u$ indicates that in year n, over a period of ℓ consecutive days ending at day k, the daily temperatures were always above a critical threshold u. On the other hand, $\mathcal{D}_\ell^{\text{avg}}(X_n)(\mathbf{s}_i, t_k) > u$ indicates that the average of the daily temperatures for the period of ℓ days ending at day t in year n exceeds level u.

Spatial functionals

To quantify the spatial extent of a heat wave, we consider for example, either the spatial average or the spatial minimum of the curves $Y(\mathbf{s}_i, t_k)$ over a neighborhood of sites $\mathcal{N}_d(\mathbf{s}_i)$ around $\mathbf{s}_i$. Specifically, we define

$$\mathcal{S}_d(Y)(\mathbf{s}_i, t_k) = \mathcal{S}_d^{\min}(Y)(\mathbf{s}_i, t_k) := \min_{\mathbf{s}\in\mathcal{N}_d(\mathbf{s}_i)} Y(\mathbf{s}, t_k) \tag{9}$$

and

$$\mathcal{S}_d(Y)(\mathbf{s}_i, t_k) = \mathcal{S}_d^{\text{avg}}(Y)(\mathbf{s}_i, t_k) := \frac{1}{|\mathcal{N}_d(\mathbf{s}_i)|} \sum_{\mathbf{s}\in\mathcal{N}_d(\mathbf{s}_i)} Y(\mathbf{s}, t_k),$$

where $|\mathcal{N}_d(\mathbf{s}_i)|$ is the number of elements in neighborhood $\mathcal{N}_d(\mathbf{s}_i)$.

One can consider various types of neighborhoods, depending on the type of available data. For example, $\mathcal{N}_d(\mathbf{s}_i)$ can denote the neighborhood of sites $\{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_m\}$ within distance d of location $\mathbf{s}_i$. With a slight abuse of notation, $\mathcal{N}_d(\mathbf{s}_i)$ can also be used to denote the neighborhood of d nearest neighbors of location $\mathbf{s}_i$ (including the location itself).

When applied to the raw temperature data, $\mathcal{S}_d^{\min}(X_n)(\mathbf{s}_i, t_k) > u$ indicates that on day k of year n, the temperature over the *entire* neighborhood $\mathcal{N}_d(\mathbf{s}_i)$ exceeds a critical threshold u. If $\mathcal{S}_d^{\text{avg}}$ is used instead, we obtain a *softer* criterion where the average temperatures in the neighborhood exceed the critical level.

Heat wave functionals

The duration and spatial extent functionals can be combined in order to obtain a variety of *heat wave functionals*. One can consider for example

$$\mathcal{H}^{(m,m)}(Y)(\mathbf{s}_i, t_k) := \mathcal{S}_d^{\min} \circ \mathcal{D}_\ell^{\min}(Y)(\mathbf{s}_i, t_k) = \min_{\mathbf{s}\in\mathcal{N}_d(\mathbf{s}_i)} \left(\min_{l=k-\ell+1,\ldots,k} Y(\mathbf{s}, t_k) \right).$$

Others include $\mathcal{H}^{(m,a)} := \mathcal{S}_d^{\min} \circ \mathcal{D}_\ell^{\text{avg}}$, $\mathcal{H}^{(a,m)} := \mathcal{S}_d^{\text{avg}} \circ \mathcal{D}_\ell^{\min}$, or in fact $\mathcal{H}^* := \mathcal{D}_\ell^{\text{avg}} \circ \mathcal{S}_d^{\min}$. Notice that the $\mathcal{S}_d^{\text{avg}} \circ \mathcal{D}_\ell^{\text{avg}} = \mathcal{D}_\ell^{\text{avg}} \circ \mathcal{S}_d^{\text{avg}}$, but $\mathcal{S}_d^{\min} \circ \mathcal{D}_\ell^{\text{avg}} \neq \mathcal{D}_\ell^{\text{avg}} \circ \mathcal{S}_d^{\min}$, and in fact all of the above examples lead to different heat wave functionals that can be used to identify (slightly and sometimes fundamentally) different heat wave events. For example, if $\mathcal{H}^{(m,m)}(X_n)(\mathbf{s}_i, t_k) > u$, then we know that there is a period of ℓ consecutive days when the temperatures over an entire d-neighborhood of site $\mathbf{s}_i$ were always *above* level u. Such an event, even if u is moderately large, can lead to rather severe socio-economic consequences. On the other hand such events may be rather rare. Therefore, *softer* heat wave functionals such as $\mathcal{H}^{(m,a)}$ and $\mathcal{H}^{(a,a)} := \mathcal{S}_d^{\text{avg}} \circ \mathcal{D}_\ell^{\text{avg}}$ are also of great practical interest. The functionals $\mathcal{H}^{(a,m)} = \mathcal{S}_d^{\text{avg}} \circ \mathcal{D}_\ell^{\min}$ or $\mathcal{H}^{(m,m)} = \mathcal{S}_d^{\min} \circ \mathcal{D}_\ell^{\min}$ involving moving minimum duration functionals are of particular interest in the context of wildfires. Indeed, a prolonged period of high temperatures leads to dry-out conditions that prepare perfect fuel for forest fires.

Heat wave magnitude and definition

The heat wave functionals defined above depend on the key parameters ℓ and d. The intensity of a heat wave depends on the threshold parameter u. These parameters can be varied with the context of the problem in order to quantify all three features of heat waves: *amplitude* (u), *duration* (ℓ), and *spatial extent* (d). Taken together, these parameters quantify the *magnitude* of a heat wave.

Additionally, for a fixed heat wave functional $\mathcal{H}$, duration ℓ, extent d, and amplitude u, a *heat wave event occurs* at location $\mathbf{s}_i$ and time t_k when $H(Y)(\mathbf{s}_i, t_k) \geq u$.

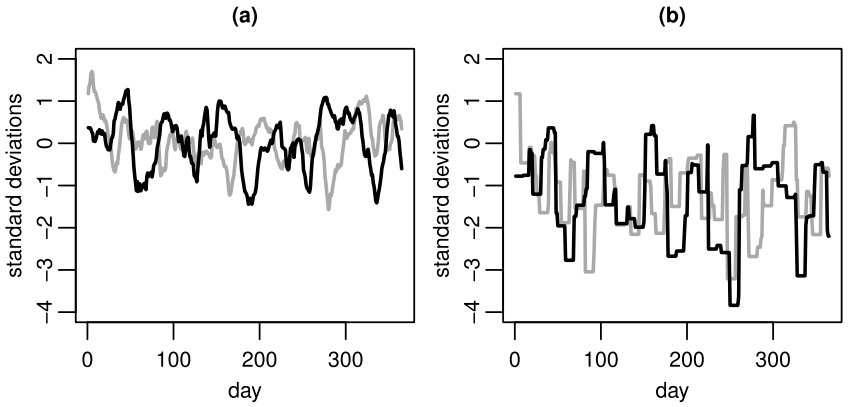

Fig. 5. Time series produced by applying duration functionals $\mathcal{D}_\ell$ to the standardized temperatures for the Wyoming site using $\ell = 10$. Time series having the same color correspond to the same year. Panel (a) shows the 10-day running means $\mathcal{D}_{10}^{\text{avg}}(Z_n)(\mathbf{s}_i, t_k)$ as a function of $k = 1, \ldots, 365$ for two different years. Panel (b) shows the 10-day running minimum, based on the $\mathcal{D}_{10}^{\min}$ functional applied to the same data.

Fig. 6. A map of the neighborhood structures for different locations using 50, 150, and 450 nearest neighbors. Each $\times$ marks a neighborhood centroid and the sequences of gray shading mark the extents of the increasing neighborhood sizes.

Illustration with NARCCAP data

Using the NARCCAP data, we now illustrate how different choices of $\mathcal{H}$ affect the results. For ease of interpretation, in the sequel, we work only with the standardized observations $Z_n(\mathbf{s}_i, t_k)$ defined by (4).

Consider the application of the $\mathcal{D}_\ell$ defined by (7) or (8) to the standardized time series $Z(\mathbf{s}_i, \cdot)$ for the site in Wyoming using a duration of $\ell = 10$ days. Thus, we are computing $\mathcal{D}_{10}(Z_n)(\mathbf{s}_i, t_k)$ for a specific location $\mathbf{s}_i$ across all times. A plot of the moving average duration loss function given in (7) for two different years is shown in Fig. 5(a). The time series measures the average amplitude over the last 10 days. The running minimum loss function given in (8) is shown in Fig. 5(b). The time series in Fig. 5(b) shows the minimum amplitude during the last ℓ days. It is seen that at this *single* location, $\mathcal{D}_{10}(Z_n)(\mathbf{s}_i, t_k) > 1$ is a rare event if (7) is used, while $\mathcal{D}_{10}(Z_n)(\mathbf{s}_i, t_k) > 0$ is a rare event if (8) is used.

We next look at plots related to the spatial extent loss function $\mathcal{S}_d^{\min}$ using $d = 50$, 150, and 450 nearest neighbors. As shown in Fig. 6, the shape of the neighborhoods changes depending on where $\mathbf{s}_i$ is located, especially near the border of the study area. We apply the spatial extent function (9) *after* the duration functionals have been applied to the standardized NARCCAP data. We consider both duration functionals (7) and (8) with $\ell = 10$. The resulting values of $\mathcal{H}(Z_{n_j})(\mathbf{s}, t_{k_j})$ (as a function of $\mathbf{s}$) for $j = 400$ are shown in Fig. 7. Recall (3) and note that $j = 400$ corresponds to day $k_j = 35$ (February 2) of year $n_j = 2$. The spatial structures of the two versions of the heat wave statistics clearly differ. In agreement with what we have seen in Fig. 5 for a single location, the heat wave statistics based on the running minima duration function are smaller than those based on the running mean duration function. This shows that extreme heat waves in the space–time domain must be defined using a different amplitude threshold u depending on the specific loss function. Furthermore, the different functionals reveal different structure of the regions under heat wave risk. As seen in Fig. 7(a), there is a region in the southwest of the United States in close proximity to Tucson, Arizona under severe heat wave risk, where the temperatures are close to $u = 1$ standard deviations above the mean for a period of 10 consecutive days. A less intense but clearly outlined region under heat wave stress extends across a rather wide area in the Alberta, Saskatchewan and British Columbia provinces of Canada along both sides of the Canadian Rocky Mountains. This observation is in line with the historical incidence of

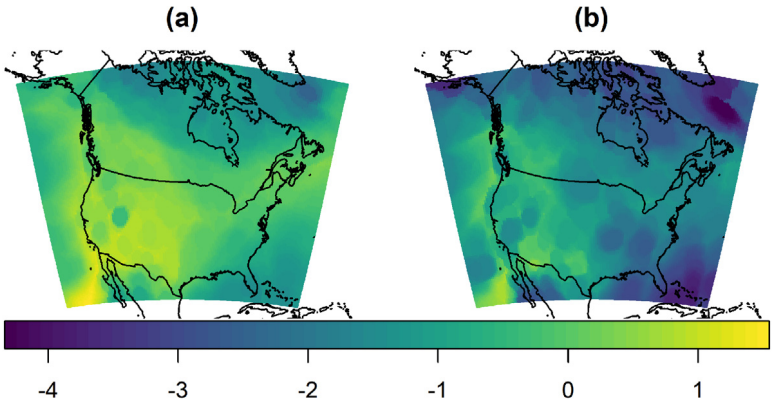

Fig. 7. Heat map of the heat wave statistic $\mathcal{H}(Z_{n_j})(\mathbf{s}, t_{k_j})$ for $j = 400$ with spatial extent $d = 50$ nearest neighbors and duration $\ell = 10$ days as a function of the spatial location $\mathbf{s}$. Panel (a) shows $\mathcal{H}^{(m,a)} = \mathcal{S}_{50}^{\min} \circ \mathcal{D}_{10}^{\text{avg}}$, while (b) shows $\mathcal{H}^{(m,m)} = \mathcal{S}_{50}^{\min} \circ \mathcal{D}_{10}^{\min}$.

wildfires in this region, which is likely to intensify in the presence of heat waves of long *durations*. We note that the maps in Fig. 7 exhibit merely the values of risk functional on a specific single day. To obtain more systematic quantification of the risk, we must be able to compute annualized probabilities of heat waves of any type. In the following section, we will develop methodology for computing probabilities of such extreme heat wave events that can be applied to general functionals $\mathcal{H}$.

3.3. The probability of a heat wave

To make the exposition more concrete, we first consider the probability of a heat wave occurring in any given calendar year. Heat wave probabilities across the entire year will be studied in Section 4.1. One can also compute probabilities of a heat wave occurring over any specified part of the year; in Section 4.2, we consider the April–May period.

Fix a location $\mathbf{s}_i$. We are interested in the probability of a heat wave of amplitude u and duration ℓ in a neighborhood $\mathcal{N}_d(\mathbf{s}_i)$ occurring during period $[a, b] \subseteq [0, 1]$ of a calendar year for a general temperature process Y. Note that $[a, b]$ is a portion of the annual calendar year. This could be the entire year, the month of April, the summer season, etc. We assume that the probability does not depend on the calendar year. Fixing the heat wave functional $\mathcal{H}$, let

$$M_n := \max_{a \le t_k \le b} \mathcal{H}(Y_n)(\mathbf{s}_i, t_k),$$

where the dependence of M_n on $\mathcal{H}$, $[a, b]$, and $\mathbf{s}_i$ is suppressed for simplicity (though clear from the context). The random variables $M_1, M_2, \ldots, M_N$ are block maxima of a weakly dependent stationary time series and are assumed to be independent and identically distributed according to a generalized extreme value (GEV) distribution, see e.g. Theorem 2.2 in Chavez-Demoulin and Davison (2012). Denote by M_0 a random variable with the same distribution as each M_n, $n = 1, 2, \ldots, N$. Then

$$\mathbb{P}(M_0 \le u) = G_{\gamma,\mu,\sigma}(u) := \exp\left\{-\left(1 + \gamma \frac{(u-\mu)}{\sigma}\right)_+^{-1/\gamma}\right\}, \tag{10}$$

with $(x)_+ = \max\{0, x\}$, and $\gamma \in \mathbb{R}$, $\mu \in \mathbb{R}$, and $\sigma > 0$, respectively, the shape, location, and scale parameters of the GEV distribution. These parameters depend on $\mathbf{s}_i$, $\mathcal{H}$, and $[a, b]$.

A heat wave with amplitude u occurs in year n if $\mathcal{H}(Y_n)(\mathbf{s}_i, t_k) > u$ for some $a \le t_k \le b$, or equivalently, if $M_n > u$. The probability

$$p(\mathbf{s}_i; \mathcal{H}, u, \ell, d, [a, b]) = P(M_n > u) = P(M_0 > u)$$

can be approximated by the tail probability of the fitted GEV distribution.

Blocks of equal size, determined by the length of $[a, b]$, must be used. What matters is that the number of elements in each block is large enough to ensure that the approximation by a GEV distribution holds reasonably well and that the number of blocks is large enough that a maximum likelihood fit to this distribution is reasonable. One can, for example, consider blocks which exclude winter months, if only probabilities of heat waves occurring during the warm season are of interest. If the time period (within a year) of interest is short, and the region under consideration climatically uniform, it is acceptable to work with the raw data $X(\mathbf{s}_i, j)$. In the following, we describe the general procedure, in which Y denotes raw, standardized, or otherwise transformed data. The condition for its applicability is that $\mathcal{H}(Y)(\mathbf{s}_i, j), j = 1, 2, \ldots, T$ be a stationary weakly dependent time series, as seen in Fig. 8.

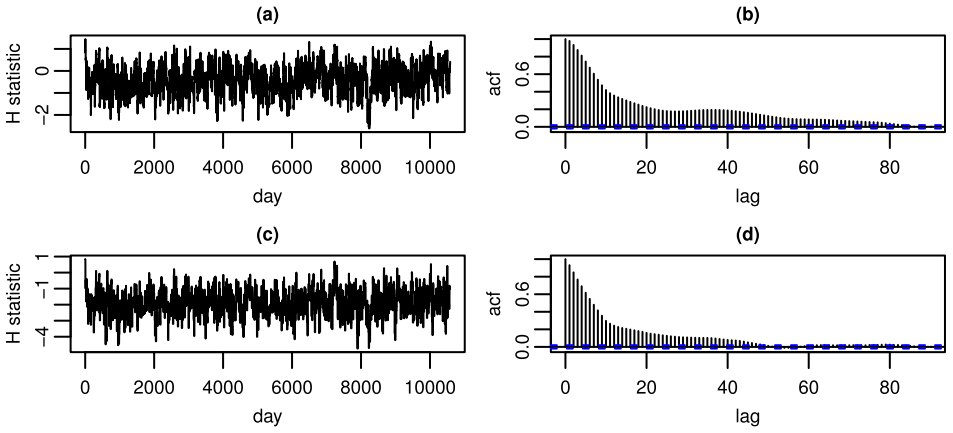

Fig. 8. A time series plot of $\mathcal{H} = \mathcal{S}_{50}^{\text{avg}} \circ \mathcal{D}_{10}^{\min}$ is shown in (a), while (b) shows the corresponding ACF plot, for the site in Wyoming. Panels (c) and (d) show the same information for $\mathcal{H} = \mathcal{S}_{50}^{\min} \circ \mathcal{D}_{10}^{\min}$.

In the general case, we want to compute the probability

$$p(\mathbf{s}_i;\, \mathcal{H}, u, \ell, d, [a, b]) := \mathbb{P}\big(\mathcal{H}(Y_n)(\mathbf{s}_i, t_k) \geq u \ \text{ for some } t_k \in [a, b]\big) \tag{11}$$

for some $[a, b] \subseteq [0, 1]$. We proceed as follows:

1. Compute the heat wave statistics
$$M_n := \max_{t_k \in [a,b]} \mathcal{H}(Y_n)(\mathbf{s}_i, t_k), \quad n = 1, 2, \ldots, N.$$
2. Fit a GEV model $G_{\gamma,\mu,\sigma}$ to the data $M_n,\ n = 1, \ldots, N$, using an appropriate method of estimation to obtain parameter estimates $\hat{\gamma}$, $\hat{\mu}$, and $\hat{\sigma}$.
3. Given a threshold u, approximate the desired heat wave probability in (11) by
$$\hat{p}(\mathbf{s}_i;\, \mathcal{H}, u, \ell, d, [a, b]) := 1 - G_{\hat{\gamma},\hat{\mu},\hat{\sigma}}(u).$$

4. Analysis of NARCCAP Data

The statistical framework we developed is now applied to climate model output from the NARCCAP data introduced in Section 2. Section 4.1 presents selected results of the application to the whole data set, while Section 4.2 focuses on the Canadian plains, which recently experienced a heat wave leading to catastrophic wildfires on a scale not previously seen. We emphasize that the maps displayed in this section do not show any cumulative probabilities for large geographical areas. The probabilities, coded by color, refer to a neighborhood centered at a given pixel. The maps are thus graphical, not inferential, tools aimed at identifying regions at risk of heat waves.

4.1. Application to the whole data set

We explore the probabilities and locations of future heat waves of various magnitudes using the NARCCAP data for different durations ℓ, and spatial extent d, while holding the amplitude threshold u fixed at 2. The amplitude $u = 2$, corresponding to two standard deviations above the mean, is chosen as a standard reference level. It allows us to focus on extreme heat waves generally not seen in historical records. It can be modified without difficulty if more or less intense heat waves are of interest. There is an interplay between the values of u, d and ℓ that can be exploited in a specific public health or agricultural context. In this section, the heat wave functional is fixed to be $\mathcal{H}^{(m,a)} := \mathcal{S}_d^{\min} \circ \mathcal{D}_\ell^{\text{avg}}$; Section 4.2 studies the effect of choosing different loss functionals. We construct heat maps of heat wave probability based on duration values $\ell = 3$, 10, and 30 days, and d nearest-neighbor neighborhoods with $d = 50$, 150, and 450. Informally, $\ell = 3$ would represent a heat wave of relatively short duration that may cause public health problems, while $\ell = 30$ would be a relatively long duration heat wave that may affect agricultural production. The spatial extent $d = 50$ would be a fairly small regional heat wave, while $d = 450$ would correspond to a heat wave affecting an area comparable to a typical US state.

Fig. 9 shows a map of the probability of a heat wave for $d = 50$ for all the duration parameters, with (a) corresponding to $\ell = 3$, (b) to $\ell = 10$, and (c) to $\ell = 30$. When $\ell = 3$, there is a surprisingly high probability of localized heat waves over locations in the Labrador Peninsula. Such short heat spells may occur with probability approaching 50%, or every other year, on average. While our EVT approximation may break down for such high probabilities, it is nevertheless obvious that such

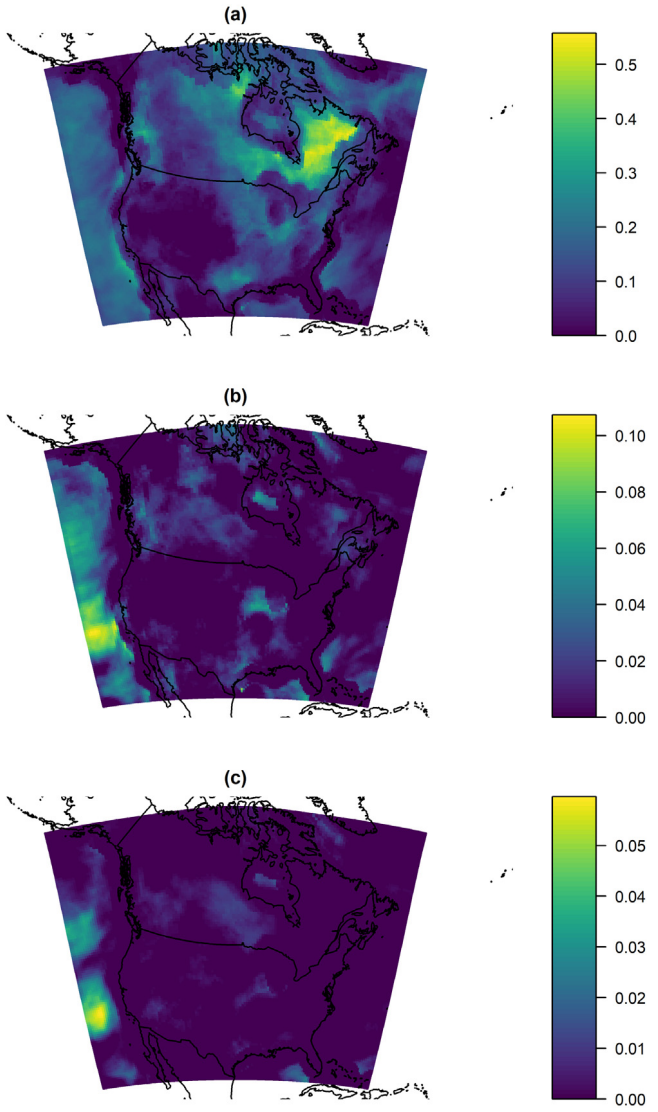

Fig. 9. Probability of a heat wave with amplitude more than 2 standard deviations above the mean for spatial extent $d = 50$ and durations of (a) $\ell = 3$, (b) $\ell = 10$, and (c) $\ell = 30$.

heat spells will be common in that part of Canada. Generally, our simple method shows that the area around the Hudson Bay will experience a high frequency of hot spells lasting a few days. There is a noticeable drop in the probability of such a heat wave around the Rocky Mountain range. The probability is also very low along the Eastern seaboard of the United States. Increasing the duration to $\ell = 10$ days, dramatically reduces the probability of a heat wave of the corresponding magnitude. The reader will note the different probability scale. Many parts of Canada once again show an increased probability of a heat wave of this magnitude, as well as parts of Iowa and Illinois, certain regions in Texas, and, most visibly, the Pacific Ocean off the Southern California coast. Increasing the duration to approximately 1 month ($\ell = 30$), causes the probability of a heat wave to drop even further; generally, throughout North America, heat waves of this magnitude will occur with probability of less than 1%, i.e. once per one hundred years, on average. Over the Canadian plains and the Canadian Rockies, this probability increases only slightly to about 1.5%. There are two patches, in Arizona and Southern Texas, with probabilities elevated to 2%–3%.

We now consider heat waves of larger spatial extent ($d = 150$) using the same durations ($\ell = 3, 10, 30$) and amplitude threshold ($u = 2$) as before. The associated probability maps are shown in Fig. 10. Comparing Figs. 9 (for $d = 50$) and 10 (for $d = 150$), we see somewhat similar spatial patterns, though the probabilities are substantially lower for $d = 150$ compared to $d = 50$. This perfectly matches our intuition, as heat waves with greater spatial extent, keeping duration and amplitude constant, should be less likely than those of smaller spatial extent. Heat waves of duration greater than $\ell = 10$ days can

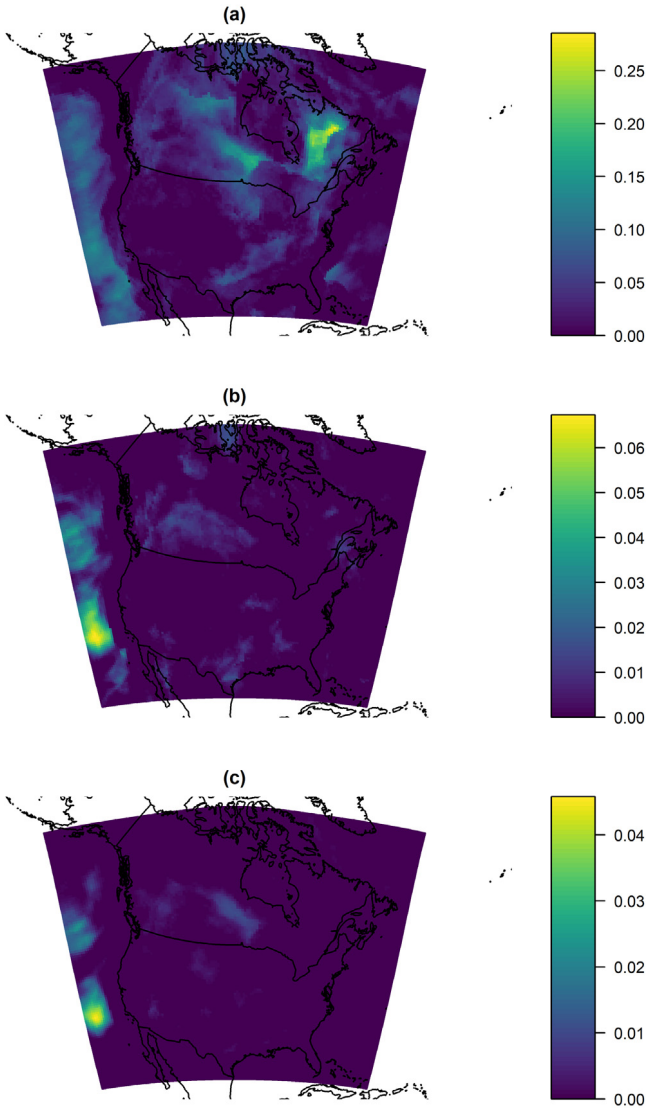

Fig. 10. Probability of a heat wave with amplitude more than 2 standard deviations above the mean for spatial extent $d = 150$ and durations of (a) $\ell = 3$, (b) $\ell = 10$, and (c) $\ell = 30$.

occur over the continental United States with a probability of less than 1%. Short heat spells of duration $\ell = 3$ days can occur over the Labrador peninsula with probability approaching 25%.

Lastly, we consider the probability of a heat wave with a spatial extent of $d = 450$ using the same duration and amplitude parameters as before. This is a relatively large spatial extent corresponding to an area of size similar to that of Colorado or Utah. The probability maps shown in Fig. 11 have a spatial structure broadly similar to that in Figs. 9 and 10 (though on different probability scales). However, there are some geographic changes in which regions are relatively more likely to experience a heat wave. Specifically, comparing panel (c) of Figs. 9–11, we see that parts of North Dakota and eastern Montana are relatively more likely to see a heat wave for $d = 450$ than for $d = 50$ or 150 compared to surrounding regions, though the absolute probability is still much lower.

4.2. Application to Canadian plains

We now proceed to analyze the probability of different magnitude heat waves (with respect to the functional $\mathcal{H}$, duration ℓ, and amplitude u) for the region surrounding Fort McMurray in Alberta, Canada. This part of Canada experienced a devastating wildfire that began on May 1, 2016. The fire necessitated the largest evacuation in Alberta history and could prove to the costliest natural disaster in Canadian history (Ramsay and Shum, 2016; Evans, 2016). The region experienced

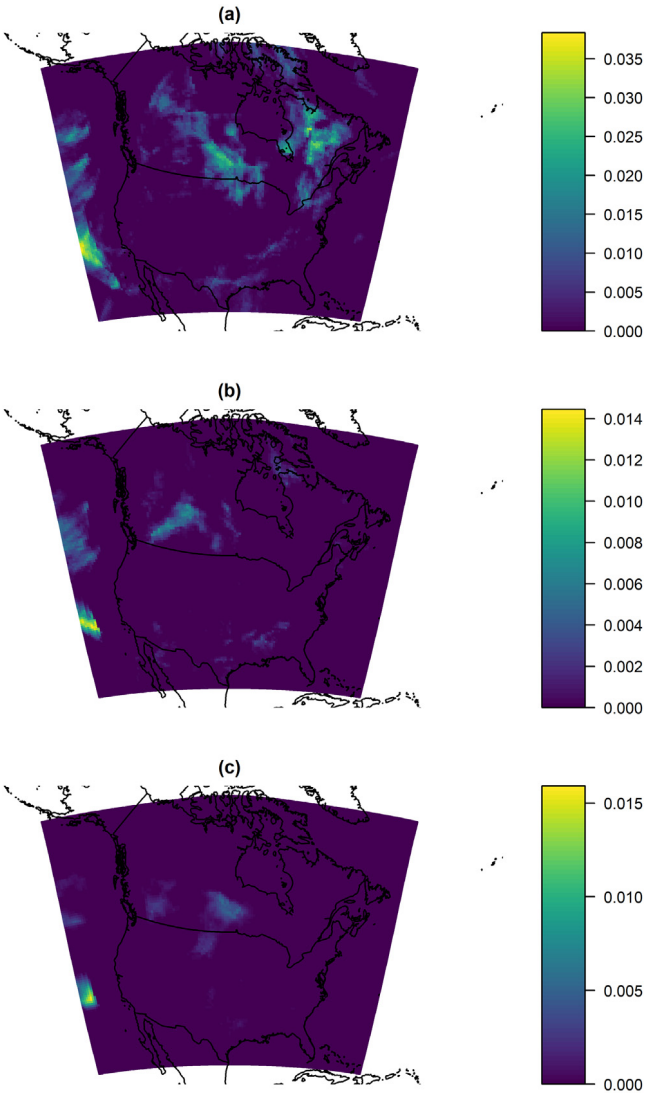

Fig. 11. Probability of a heat wave with amplitude more than 2 standard deviations above the mean for spatial extent $d = 450$ and durations of (a) $\ell = 3$, (b) $\ell = 10$, and (c) $\ell = 30$.

unusually hot temperatures and low humidity prior to and during the fire, which is believed to have aided its spread. In what follows, we examine the temperature conditions immediately prior to the fire and estimate the probability of seeing similar conditions in a given year using the same NARCCAP data analyzed in Section 4.2. A more complete analysis might also use humidity, precipitation and wind data.

We first consider the temperature conditions in Fort McMurray immediately prior to the fire. Maximum daily temperature (°C) was obtained for the weeks prior to the fire (Accuweather.com, 2016). Three different duration loss functions (average, median, and minimum) were crossed with three different duration lengths (1 week, 2 weeks, 3 weeks) to quantify the temperature conditions in various ways. More specifically, duration functionals $\mathcal{D}_\ell^{\text{avg}}$ (7), $\mathcal{D}_\ell^{\text{min}}$ (8), and $\mathcal{D}_\ell^{\text{med}}$ defined by

$$\mathcal{D}_\ell^{\text{med}}(X)(\mathbf{s}_i, t_k) := \underset{l=k-\ell+1,\ldots,k}{\text{median}} X(\mathbf{s}_i, t_l), \tag{12}$$

were applied to raw temperatures with durations $\ell = 7$, 14, and 21 days. The values produced by the duration functionals on the day before the fire began (April 30, 2016) are shown in Table 1. Note that a spatial functional ($\mathcal{S}$) was not used in calculating these values, as only temperatures at Fort McMurray were considered. Since these values quantify, in some sense, the temperature conditions immediately prior to the fire, they were used as amplitude thresholds of interest in the analysis that follows. While we argued that standardized temperatures are more appropriate for quantifying heat waves, our time

Table 1
Temperatures (°C) for April 30, 2016 produced by different duration functions for different duration lengths when applied the maximum daily temperature for Fort McMurray, Alberta in April 2016.

Duration	Duration function		
	Average	Median	Minimum
One week	15.0	15.0	8.0
Two weeks	16.1	15.0	8.0
Three weeks	13.6	13.0	0.0

frame (two months) and location are relatively small, so we work with the raw NARCCAP temperatures (after transforming them to degrees Celsius for compatibility).

We now examine the probability of seeing a heat wave with the conditions given in Table 1. Specifically, we consider three combinations of spatial extent and duration functions: average–average, median–median, and minimum–minimum. For clarity, define

$$\mathcal{S}_d^{\text{med}}(Y)(\mathbf{s}_i, t_k) := \underset{\mathbf{s}\in\mathcal{N}_d(\mathbf{s}_i)}{\text{median}}\, Y(\mathbf{s}, t_k). \tag{13}$$

We considered heat wave functionals $\mathcal{H} = \mathcal{S}_d^{\text{avg}} \circ \mathcal{D}_\ell^{\text{avg}}$, $\mathcal{H} = \mathcal{S}_d^{\text{med}} \circ \mathcal{D}_\ell^{\text{med}}$, and $\mathcal{H} = \mathcal{S}_d^{\text{min}} \circ \mathcal{D}_\ell^{\text{min}}$. We fixed the spatial extent at $d = 50$ nearest neighbors, while for each heat wave functional, we considered heat waves lasting $\ell = 7$, 14, and 21 days. We used temperatures in April and May only, as this period includes the start of the Fort McMurray fire as its center.

We examine the probability of seeing a heat wave with characteristics similar to those before the Fort McMurray fire during the time period 2041–2069, as predicted by the NARCCAP data described in Section 2. Using the parameters described in the previous paragraph, heat maps of the associated probabilities are shown in Fig. 12. For $\ell = 7$ days, the probability of seeing similar temperature conditions near Fort McMurray is about 87% for the mean–mean heat wave functional, close to 89% for the median–median heat wave functional, and near 75% for the min–min functional. Increasing the heat wave duration to $\ell = 14$ days, the probability of seeing similar temperature conditions is close to 46% for the mean–mean functional, 67% for the median–median heat wave functional, and about 30% for the min–min functional. Perhaps surprisingly, for $\ell = 21$, this is about 68% for the mean–mean functional, 76% for the median–median heat wave functional, and 92% for the min–min functional. This rise in heat wave probability compared to $\ell = 14$ occurs because the amplitude thresholds used for the two calculations differ.

In summary, the likelihood of seeing temperature conditions similar to those immediately before the May 2016 Fort McMurray fire is fairly large when looking at heat waves based on average, median, or minimum temperature. The first two functionals are more relevant for predicting conditions conducive to wildfires as they indicate high average (median) temperatures over a long period and a large area. The min–min functional is large if the minimum temperature over an area and over a period of time never drops below a certain threshold. One very cold day somewhere in the region of interest is not likely to reduce the risk of a wildfire. The broad conclusion is that conditions conducive to the May 2016 Fort McMurray area wildfire are likely to occur in the mid 21st century with relatively high frequency based on the temperature data from the NARCCAP program.

5. Assessment of the methodology

It this section, we validate the methodology and conclusions of the previous sections using several simulation studies and sensitivity analysis.

The uncertainty of the estimated parameters can impact the computed probability of an extreme event; it is useful to quantify the size of this impact. We consider the Wyoming site used before. We created heat wave statistics using four composition functions: $\mathcal{S}_d^{\text{min}} \circ \mathcal{D}_\ell^{\text{avg}}$, $\mathcal{S}_d^{\text{avg}} \circ \mathcal{D}_\ell^{\text{avg}}$, $\mathcal{S}_d^{\text{min}} \circ \mathcal{D}_\ell^{\text{min}}$, and $\mathcal{S}_d^{\text{med}} \circ \mathcal{D}_\ell^{\text{med}}$ with $\ell = 10$ days and smoothing over neighbors within $d = 80$ km of the site. Annual block maxima of the statistics for the site were taken, and then maximum likelihood estimation was used to estimate the parameters for a GEV model fit to that set of maxima. The probability of a heat wave with amplitude more than $u = 1.75$ and $u = 2$ (standard deviations), for the four functionals was computed using the estimated parameters, as well as various combinations of the lower and upper bounds of 95% confidence intervals for each parameter (based on a normal approximation). The associated upper-tail heat wave probabilities, $\hat{p}$, are shown in Table 2 for $\mathcal{S}_d^{\text{min}} \circ \mathcal{D}_\ell^{\text{avg}}$ with $u = 1.75$ Complete results are provided in the Supplementary Material. It is seen that the probability is relatively unaffected by the different combination of parameter values, unless there are large relative changes in both the scale location and shape parameters. In general, the probabilities fluctuate more for smaller values of u, and much less for the larger values of u, which correspond to the extreme events of interest.

Another question of interest is how well the proposed methodology approximates the desired exceedance probabilities. We consider an experiment to assess this. First, we assume that the block maxima M_n, $n = 1, 2, \ldots, 29$, have the standard Gumbel distribution, i.e., $P(M_n \le x) = \exp\{-e^{-x}\}$. For an *upper tail* probability p, we can compute u_p such that

$$1 - \exp\{-e^{-u_p}\} = p,$$

i.e. $u_p = -\ln\{-\ln(1-p)\}$. For fixed p, we then execute the following simulation experiment:

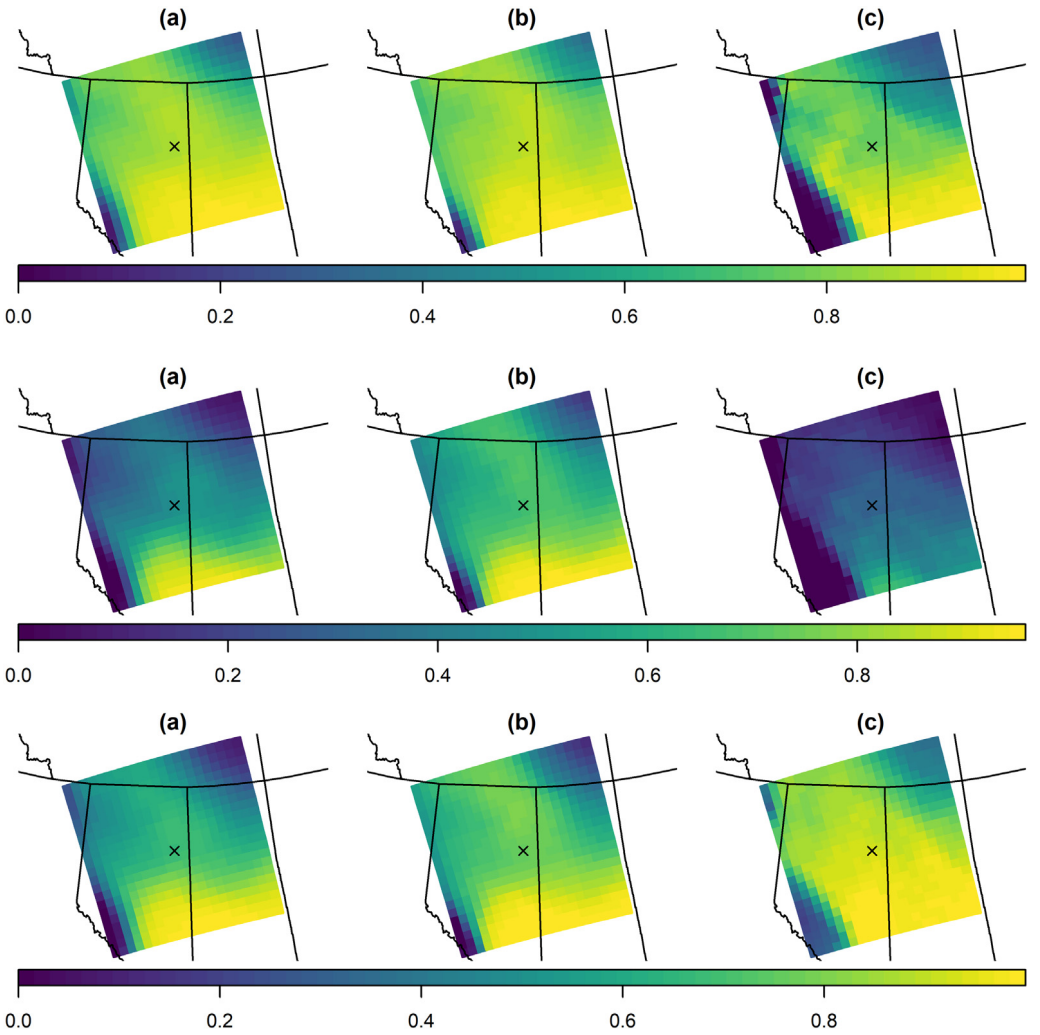

Fig. 12. Probability of a heat wave in the region surrounding Fort McMurray using conditions similar to those before the May 2016 fire. The heat wave functional for (a) is $\mathcal{H} = \mathcal{S}_d^{\text{avg}} \circ \mathcal{D}_\ell^{\text{avg}}$, (b) is $\mathcal{H} = \mathcal{S}_d^{\text{med}} \circ \mathcal{D}_\ell^{\text{med}}$, and (c) is $\mathcal{H} = \mathcal{S}_d^{\text{min}} \circ \mathcal{D}_\ell^{\text{min}}$. The duration is, from top to bottom, 7, 14, 21 days. The spatial extent is $d = 50$. The amplitude threshold used in (a) is 15.0 °C, in (b) is 15.0 °C, and in (c) is 8.0 °C. Fort McMurray is shown by the ×.

Table 2
Table of the estimated heat wave probabilities for the Wyoming site, using various combinations of parameter estimates and bounds from 95% confidence intervals, for $\mathcal{S}_d^{\text{min}} \circ \mathcal{D}_\ell^{\text{avg}}$ with $u = 1.75$.

Location	Scale	Shape	$\hat{p}$
0.92	0.25	−0.36	0.0000
1.01	0.18	−0.36	0.0000
1.01	0.25	−0.58	0.0000
1.01	0.25	−0.36	0.0000
0.92	0.18	−0.58	0.0000
1.11	0.25	−0.36	0.0005
1.01	0.32	−0.36	0.0061
1.01	0.25	−0.13	0.0212
1.11	0.32	−0.13	0.0895

1. Generate independent, identically distributed standard Gumbel M_n, $n = 1, 2, \ldots, 29$.
2. Fit a GEV model to the sample using maximum likelihood estimation to estimate model parameters $\hat{\gamma}$, $\hat{\mu}$, and $\hat{\sigma}$.
3. Compute $\hat{p} = 1 - G_{\hat{\gamma},\hat{\mu},\hat{\sigma}}(u_p)$.

Table 3
Simulation experiment results comparing the estimated upper-tail probability to the true upper-tail probability.

p	0.5	0.25	0.1	0.01	0.001
$\bar{\hat{p}}$	0.50	0.24	0.09	0.01	0.00
se($\bar{\hat{p}}$)	0.09	0.07	0.04	0.01	0.00

Fig. 13. A heat map of elevation across the NARCCAP study area (lighter colors indicate higher elevation. The ×s mark the location of Saint Louis (Missouri), District of Columbia, Vancouver (British Columbia), Fort McMurray (Alberta), Houston (Texas), Iqaluit (Nunavut), and the Wyoming site.

4. Repeat steps 1–3 100 times, then compute $\bar{\hat{p}}$, the average of $\hat{p}$ over the 100 samples, and se($\bar{\hat{p}}$), the estimated standard error of $\bar{\hat{p}}$.

The results, shown in Table 3, show that $\bar{\hat{p}}$ is close to the true value of p, with an associated standard error that decreases as the probability p becomes more extreme.

We next perform a similar simulation study, but rather than using the Gumbel distribution, we use estimated GEV distributions at seven locations spread across the study area. Specifically, we consider the sites Saint Louis (Missouri), District of Columbia, Vancouver (British Columbia), Fort McMurray (Alberta), Houston (Texas), Iqaluit (Nunavut), and the Wyoming site previously discussed. The locations of these sites are shown on a heat map of elevation in Fig. 13. Heat wave statistics for the NARCCAP data were computed using four functionals: $\mathcal{S}_d^{\min} \circ \mathcal{D}_\ell^{\text{avg}}$, $\mathcal{S}_d^{\text{avg}} \circ \mathcal{D}_\ell^{\text{avg}}$, $\mathcal{S}_d^{\min} \circ \mathcal{D}_\ell^{\min}$, and $\mathcal{S}_d^{\text{avg}} \circ \mathcal{D}_\ell^{\max}$, with $\ell = 10$ days and smoothing over neighbors within $d = 240$ km of the centroid. We then consider the estimated GEV model for the grid cell whose centroid was closest to each of the seven locations. For those grid cells, we computed $\hat{p} = 1 - G_{\hat{\gamma},\hat{\mu},\hat{\sigma}}(u)$ using the respective GEV model parameters estimated for each grid cell and the relevant amplitude u. We then:

1. Generated i.i.d. M_n, $n = 1, 2, \ldots, 29$, from a GEV($\hat{\gamma}$, $\hat{\mu}$, $\hat{\sigma}$) distribution.
2. Fit a GEV model to the sample from 1 using maximum likelihood estimation to estimate model parameters by $\tilde{\gamma}$, $\tilde{\mu}$, and $\tilde{\sigma}$.
3. Compute $\tilde{p} = 1 - G_{\tilde{\gamma},\tilde{\mu},\tilde{\sigma}}(t)$.
4. Repeat steps 1–3 100 times, then compute $\bar{\tilde{p}}$, the average of $\tilde{p}$ over the 100 samples, and se($\bar{\tilde{p}}$), the estimated standard error of $\bar{\tilde{p}}$.

The amplitude $u = 1.5$ was used for $\mathcal{S}_d^{\min} \circ \mathcal{D}_\ell^{\text{avg}}$ and $\mathcal{S}_d^{\text{avg}} \circ \mathcal{D}_\ell^{\text{avg}}$, $u = 1.25$ for $\mathcal{S}_d^{\min} \circ \mathcal{D}_\ell^{\min}$, and $u = 2.5$ for $\mathcal{S}_d^{\text{avg}} \circ \mathcal{D}_\ell^{\max}$. The results are shown in Table 4. In all cases, the mean of the sampling distribution ($\bar{\tilde{p}}$) was quite close to the probability ($\hat{p}$) being estimated.

An important aspect of this analysis is verifying the dependence assumptions of the heat wave statistics over time. Specifically, we require that the autocorrelation among heat waves statistics becomes small as the lag between them increases. A time series plot of the autocorrelation function as a function of lag distance for all locations simultaneously is not helpful, as there are over 16,000 lines to plot. Instead, we consider a heat map of the autocorrelation function across the NARCCAP domain for several lags. The autocorrelation function will drop differently for each location as a function of lag, but a heat map should show the autocorrelation steadily dropping as lag distance increases. The autocorrelation heat map of the $\mathcal{H} = \mathcal{S}_d^{\min} \circ \mathcal{D}_\ell^{\text{avg}}$, with $\ell = 10$ days and smoothing over $d = 50$ nearest neighbors, is shown for lags of (a) 10, (b) 20, (c) 30, (d) 40, (e) 50, and (f) 60 days in Fig. 14. The autocorrelations decay dramatically over the land masses of North America, with a much slower decay over bodies of water. Practically, this is not a problem in our analysis since inference is mainly desired over land, where residents are more likely to be affected by extreme events.

Lastly, we considered the effect of a (small) trend in the raw data on the estimated heat wave probabilities. In the context of the NARCCAP data, if the temperatures have a small annual increase (on average), how does this affect the heat wave

Table 4
Estimated mean and standard deviation of the sampling distribution of an estimated probability using synthetic data generated from GEV model with estimated parameter values for several sites in the NARCCAP study area. Location abbreviations refer to Saint Louis, District of Columbia, Vancouver, Fort McMurray, Houston, Iqaluit and Wyoming. The heat wave statistics for which the GEV distributions were estimated are provided on the left side of the table, along with the amplitude u for which the upper-tail probability is estimated.

		S.L.	D.C.	Van.	F.M.	Hou.	Iq.	Wy.
$\mathcal{S}_d^{\min} \circ \mathcal{D}_\ell^{\text{avg}}, u = 1.5$	$\hat{p}$	0.105	0.013	0.117	0.100	0.057	0.000	0.000
	$\bar{\hat{p}}$	0.093	0.009	0.110	0.098	0.047	0.000	0.001
	$\text{se}(\bar{\hat{p}})$	0.043	0.011	0.054	0.040	0.034	0.000	0.004
$\mathcal{S}_d^{\text{avg}} \circ \mathcal{D}_\ell^{\text{avg}}, u = 1.5$	$\hat{p}$	0.434	0.255	0.555	0.214	0.309	0.374	0.114
	$\bar{\hat{p}}$	0.432	0.269	0.563	0.210	0.298	0.377	0.103
	$\text{se}(\bar{\hat{p}})$	0.084	0.075	0.090	0.058	0.075	0.082	0.048
$\mathcal{S}_d^{\min} \circ \mathcal{D}_\ell^{\min}, u = 1.25$	$\hat{p}$	0.000	0.000	0.000	0.022	0.005	0.002	0.002
	$\bar{\hat{p}}$	0.001	0.001	0.002	0.024	0.007	0.004	0.006
	$\text{se}(\bar{\hat{p}})$	0.001	0.002	0.004	0.018	0.010	0.008	0.008
$\mathcal{S}_d^{\text{avg}} \circ \mathcal{D}_\ell^{\max}, u = 2.5$	$\hat{p}$	0.454	0.453	0.510	0.563	0.442	0.815	0.003
	$\bar{\hat{p}}$	0.455	0.476	0.518	0.579	0.442	0.812	0.006
	$\text{se}(\bar{\hat{p}})$	0.085	0.082	0.092	0.071	0.083	0.056	0.009

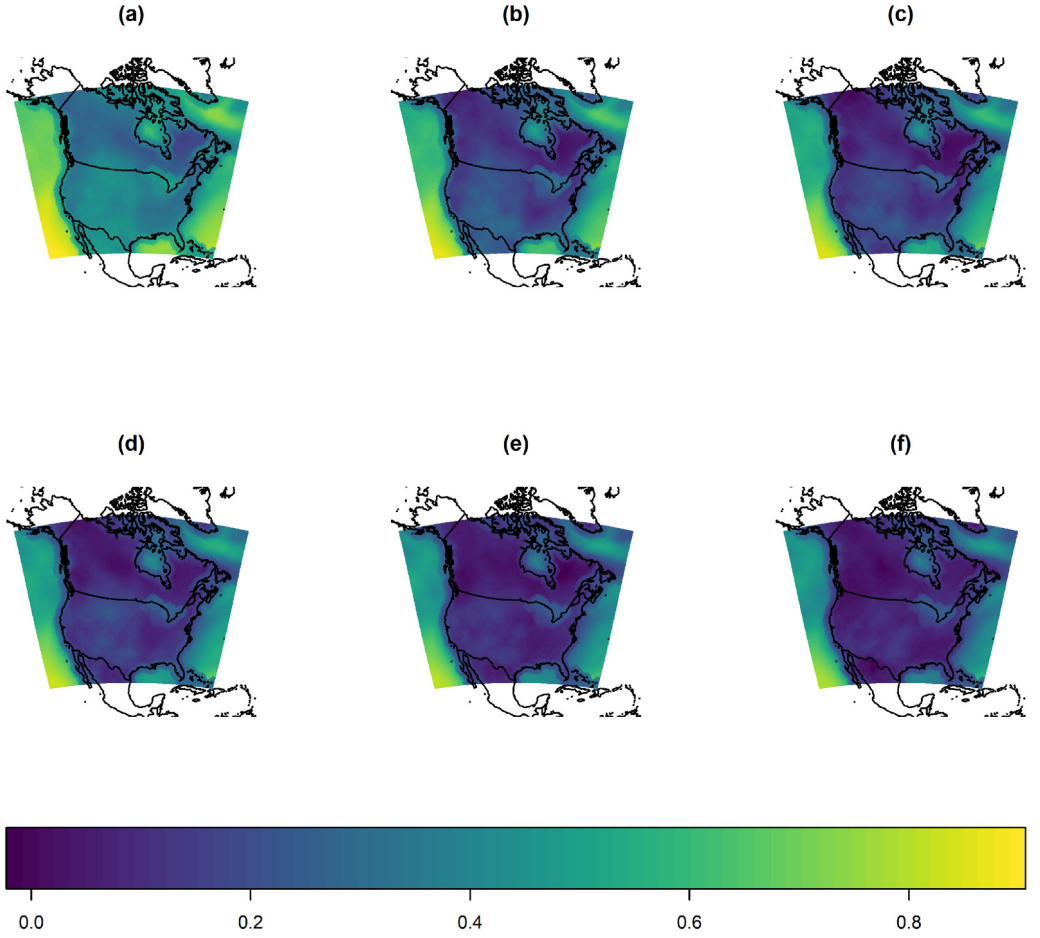

Fig. 14. A heat map of the autocorrelation function of $\mathcal{H} = \mathcal{S}_d^{\min} \circ \mathcal{D}_\ell^{\text{avg}}$ for several lags, with $\ell = 10$ days and smoothing over $d = 50$ nearest neighbors. The autocorrelation function is shown for lags of (a) 10, (b) 20, (c) 30, (d) 40, (e) 50, and (f) 60 days.

probabilities? We focused on the data used in the prior Fort McMurray analysis. We proceeded with analysis in a similar fashion, with one small change. Specifically, the raw data were "detrended" before analysis. A simple linear regression model was fit to the raw temperature time series at each location. The residuals of the fitted model were obtained, and then the fitted value for the first day of temperatures was added to the residuals (to put the detrended temperatures on the same

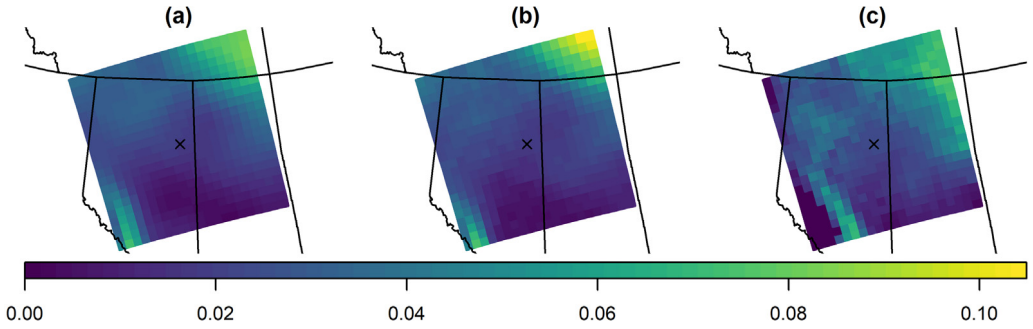

Fig. 15. Difference in the probability of a heat wave in the region surrounding Fort McMurray using conditions similar to those before the May 2016 fire, using both the original raw data and detrended data. The heat wave functional for (a) is $\mathcal{H} = \mathcal{S}_d^{\text{avg}} \circ \mathcal{D}_\ell^{\text{avg}}$, (b) is $\mathcal{H} = \mathcal{S}_d^{\text{med}} \circ \mathcal{D}_\ell^{\text{med}}$, and (c) is $\mathcal{H} = \mathcal{S}_d^{\text{min}} \circ \mathcal{D}_\ell^{\text{min}}$. The duration is 7 days and the spatial extent is $d = 50$ in each panel. The amplitude threshold used in (a) is 15.0 °C, in (b) is 15.0 °C, and in (c) is 8.0 °C. Fort McMurray is shown by the $\times$.

scale as the original data). The goal of this adjustment was to remove any small trend in the data before analysis. We then considered heat wave functionals $\mathcal{H} = \mathcal{S}_d^{\text{avg}} \circ \mathcal{D}_\ell^{\text{avg}}$, $\mathcal{H} = \mathcal{S}_d^{\text{med}} \circ \mathcal{D}_\ell^{\text{med}}$, and $\mathcal{H} = \mathcal{S}_d^{\text{min}} \circ \mathcal{D}_\ell^{\text{min}}$ with a fixed spatial extent of $d = 50$ nearest neighbors and heat waves of duration $\ell = 7$ days. After computing the block maxima of the April heat wave statistics and fitting GEV models to the block maxima, we proceeded to estimate the probability of a heat wave of amplitude greater than the values in the first row of Table 1 for each respective heat wave functional, as done in the previous analysis. We compared the probability results from the original analysis to the results from the detrended data by creating a heat map of the difference between the estimated probability from the original analysis and the estimated probability from the detrended data. The results are shown in Fig. 15. There is generally only a small difference in the probabilities. Detrending the data results in lower temperatures, so the heat wave probabilities for the detrended data are slightly smaller than for the original data.

6. Summary and outlook

Combining methods of FDA and EVT, we proposed a paradigm for computing probabilities of heat waves with specified characteristics. These probabilities can be computed once the spatial region, duration and intensity, together with a loss function have been determined. These will depend on a specific economic or climate research question, but the general methodology is the same. A specific application might be as follows. In an agricultural region that might grow a certain crop, a specific type of a heat wave might be very harmful to that crop. We can compute the probability of such a heat wave (per year) thus enhancing informed decisions about the selection of future crops.

Since we aimed at proposing a general statistical framework, we illustrated our approach with a broad application to all locations in North America and several spatial extents and durations. We also focused on a smaller region in Canada. We showed what kind of information can be obtained by drawing maps $\mathbf{s} \mapsto P(\mathbf{s})$, where $\mathbf{s}$ is a location and $P(\mathbf{s})$ is a probability of a specific heat wave centered at $\mathbf{s}$. The maps show how these probabilities change in space for different types of heat waves. They are graphical, not inferential, tools aimed at showing interesting patterns. Probabilities can be attached only to a specific region defined by its center and the neighborhood of this center.

The approach advocated in the paper in not restricted to the quantification and computation of the probabilities of heat waves. It can be applied to other extreme weather events that are characterized by temporal duration and spatial extent. For example to computing the probability of a drought. A drought is characterized by unusually low precipitation over a region for an extended period of time. The duration parameter ℓ could thus correspond to 2–3 months. More details are given in the Supplemental Material.

Going beyond weather data, spatially indexed functional data naturally occur as hyperspectral data in remote sensing, see e.g. Zullo et al. (2018) and references therein. It is conceivable that a combination of EVT and FDA methods might, for example, lead to methodology for detecting the spread of invasive species. Another potential direction might be prediction of extreme pollution events. FDA techniques have been applied to pollution data, but with different goals, see e.g. Hörmann et al. (2018) and references therein. It is hoped that this paper will stimulate research that combines approaches of EVT and FDA to solve many practical questions.

Interested researchers may find the NARCCAP data analyzed in this paper at the permanent URL https://dx.doi.org/10.17632/jz553h7ytw.1. R code for producing the graphics and results are available at runmycode.org.

Acknowledgments

The authors have been supported by NSF grant "FRG: Collaborative Research: Extreme Value Theory for Spatially Indexed Functional Data" (1463642 CU Denver, 1462067 CSU, 1462368 Michigan). J. French was also supported under NIH grant R01 CA157528.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.csda.2018.07.004.

References

Accuweather.com (2016). URL http://www.accuweather.com/en/ca/fort-mcmurray/t9h/april-weather/51933 Accessed on July 15, 2016.
Beniston, M., 2004. The 2003 heat wave in Europe: A shape of things to come? An analysis based on Swiss climatological data and model simulations. Geophys. Res. Lett. 31 (2).
Burt, C.C., 2016. Possible New Continental Heat Record for Antarctica. URL https://www.wunderground.com/blog/weatherhistorian/comment.html?entrynum=323#commenttop Accessed on May 4, 2016.
Caya, D., Laprise, R., 1999. A semi-implicit semi-Lagrangian regional climate model: The Canadian RCM. Mon. Weather Rev. 127 (3), 341–362.
Chavez-Demoulin, V., Davison, A.C., 2012. Modelling time series Extremes. REVSTAT 10, 109–133.
Collins, M., Booth, B.B., Harris, G.R., Murphy, J.M., Sexton, D.M., Webb, M.J., 2006. Towards quantifying uncertainty in transient climate change. Clim. Dynam. 27 (2–3), 127–147. http://dx.doi.org/10.1007/s00382-006-0121-0.
Evans, P., 2016. Fort McMurray fire could cost insurers 9B, BMO predicts. CBC News, May 5, 2016. URL http://www.cbc.ca/news/business/fort-mcmurray-insurance-cost-1.3568113. Accessed on July 15, 2016.
Gershunov, A., Cayan, D.R., Iacobellis, S.F., 2009. The Great 2006 Heat Wave over California and Nevada: Signal of an Increasing Trend. J. Clim. 22 (23), 6181–6203.
Hajat, S., Kovats, R.S., Atkinson, R.W., Haines, A., 2002. Impact of hot temperatures on death in London: A time series approach. J. Epidemiol. Community Health 56, 367–372.
Hörmann, S., Kokoszka, P., Nisol, G., 2018. Testing for periodicity in functional time series. Ann. Statist. 00, Forthcoming, 0000–0000.
Mearns, L., et al. (2007, updated 2014). The North American Regional Climate Change Assessment Program dataset, National Center for Atmospheric Research Earth System Grid data portal, Boulder, CO http://dx.doi.org/10.5065/D6RN35ST.
Mearns, L.O., Gutowski, W., Jones, R., Leung, R., McGinnis, S., Nunes, A., Qian, Y., 2009. A regional climate change assessment program for North America. EOS, Trans. Amer. Geophys. Union 90 (36), 311–311.
Ramsay, C., Shum, D., 2016. 'Ocean of fire' destroys 2,400 structures but 85% of Fort McMurray still stands. Global News, May 9, 2016. URL http://globalnews.ca/news/2688553/notley-in-fort-mcmurray-monday-to-survey-wildfire-damage/ Accessed on July 15, 2016.
Zullo, A., Fauvel, M., Ferraty, F., 2018. Experimental comparison of functional and multivariate spectral-based supervised classification methods in hyperspectral image. J. Appl. Stat. Published online.